WEST BENGAL STATE UNIVERSITY

Sentiment Analysis From
Bengali Text

Suny Dutta

A Dissertation
submitted to the Faculty of Graduate Studies
in partial fulfillment of the requirements
for the Degree of
M.Sc. (Computer Science)

Post Graduate Program in Computer Science
West Bengal State University
West Bengal, India

July 2022

Acknowledgments

I hereby declare that the dissertation **Sentiment Analysis from Bengali Text** submitted by me to the **Department of Computer Science**, **West Bengal State University** in partial fulfilment for the award of **Master of Science** in **Computer Science** is a bonafide record of the work carried out by me under the supervision of **Prof. Kaushik Roy**. I further declare that the work reported in this dissertation, has not been submitted and will not be submitted, either in part or in full, for the award of any other degree or diploma of this institute or of any other institute or University.

Date: July 13, 2022

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

Sentiment analysis, also known as opinion mining, is a field within Natural Language Processing (NLP) that consists in automatically identifying the sentiment of a text, often in categories like negative, neutral and positive. It has become recently popular in both research and business due to the large and increasing amount of opinionated text from Internet users, such as social media platforms and reviews. Knowing how users feel or think about a certain brand, product, idea or topic is a valuable source of information for companies, organizations and researchers, but it can be a challenging task. Natural language often contains ambiguity and figurative expressions that make the automated extraction of information in general very challenging.

Traditional sentiment analysis focuses on classifying the overall sentiment of a text without specifying what the sentiment is about. This may not be enough if the text is simultaneously referring to different topics or entities (also known as aspects), possibly expressing different sentiments towards different aspects. Identifying sentiments associated to specific aspects in a text is a more complex task known as aspect-based sentiment analysis (ABSA). ABSA has been a research topic that gained traction during SemEval-2014 Workshop 2, where it was first introduced as Task 4. The task was, given a text about restaurants or laptops, to identify the aspects of the given topic and predict the sentiment polarity for each aspect that was identified.

There have been many recent developments in the field of pre-trained NLP models, for example ELMo [1], Universal Language Model Fine-tuning (ULMfit) [2] and BERT [3]. These NLP models are pre-trained on large amounts of unannotated text. Their use has shown to allow better performance with a reduced requirement for labeled data and also much faster training. At SemEval-2016, there were no submissions that used such pre-trained NLP model as a base for the ABSA tasks.

1.2 Aim

The Primary Aim of this work is to study the BERT model when fine-tuning BERT on Text/Reviews from multi-domain dataset.

1.3 Project Outline

The study begins with a Related Work in Chapter 2. Chapter 3 explores the paradigms of deep learning, sentiment analysis, and social media analytics. Chapter 4 describes the problem, methodology used, and the data used in this study. Chapter 5 displays the Experimental results. The summary part concludes the work of this thesis in Chapter 6 where we discussed the problem, outcome of this study and Future Work.

Chapter 2

Related Work

Sentiment analysis of customer reviews has become a hot research area. Many scholars and researchers have done their research on sentiment analysis where different emotions are categorized and various machine learning models were implemented after applying different NLP techniques. While there are lots of works and NLP tools for the English language, scarcity of Bengali annotated datasets and no proper NLP tools for Bangla text preprocessing are some main restrictions on more research in this discipline.

In [4], have developed four models with the hybrid of Convolutional Neural Network (CNN) and Long Short term memory (LSTM) with various Word Embeddings including Embedding Layer, Word2Vec, Global Vectors (Glove), and Continuous Bag of Words (CBOW) to detect emotion from Bangla texts (words, sentences). The emotions they tried to detect from text are happiness, anger, and sadness. And in result, the best model integrating Word2Vec embedding layer with a hybrid of CNN-LSTMdetected emotions from raw textual data with an accuracy of 90.49% and F1 score of 92.83%.

In [5] intends to deploy Long Short Term Memory (LSTM) Deep Recurrent Network for sentiment analysis on Bangla Text as it is developed to avoid long-term dependency. Their work show the effects of hyper-parameter tuning and the ways it can be helpful for sentiment analysis on the dataset. The goal of this study is to establish a sentiment classification framework to analyze the performance of different deep learning models with a variety of parameter calibration combinations. The proposed LSTM model with advanced layers achieves better performance on resolving sentiment polarity for aimed entities with an accuracy of 94

In [6], created dataset by collecting data from different social media platforms and websites and manually handcrafted and labled them. Their work presents how the attention mechanism could be incorporated effectively and efficiently in analyzing the Bangla sentiment or opinion.

[7] aims at detecting multi-class emotions from Bangla text using Multinomial Naïve Bayes (NB) classifier along with various features such as stemmer, parts-of-speech (POS) tagger, n-grams, term frequency-inverse document frequency (tf-idf). Their final model was able to classify the text into three emotion classes (happy, sad and angry) with an overall accuracy of 78.6

In the paper, [8] have developed a polarity detection system on textual movie reviews in Bangla by using two popular machine learning algorithms named Naive Bayes and Support Vector Machines and provided a comparative results where SVM performed slightly better than NB by considering stemmed unigram as feature with an excellent precision of 0.86.

In this paper, [9] scrape 1000 positive and 1000 negative reviews on automobile and store them in WEKA attribute-relation file format. Training was done on 80% of the data and rest of it was used for testing purpose which was done using different models and results were analyzed in each case. The results showed that Multinomial Naive Bayes outperformed Bagging, Deep Neural Network, Decision Tree, Random Forest, AdaBoost, k-NN and SVM Classifiers in terms of more accuracy, precision, recall and F-measure.

In the paper, Khondoker Ittehadul Islam et. al. 2020 [10] presents manually tagged 2-class and 3-class SA datasets in Bengali. They also demonstrate that the multi-lingual BERT model with relevant extensions can be trained via the approach of transfer learning over those novel datasets to improve the state-of-the-art performance in sentiment classification tasks. This deep learning model achieves an accuracy of 71% for 2-class sentiment classification compared to the current state-of-the-art accuracy of 68%. They also present the very first Bengali SA classifier for the 3-class manually tagged dataset, and proposed model achieves an accuracy of 60%.

Shahrukh Khan et. al. 2022 [11] explores how they can effectively use deep neural networks in transfer learning and joint dual input learning settings to effectively classify sentiments and detect hate speech in Hindi and Bengali data. They start by training Word2Vec word embeddings for Hindi HASOC dataset and Bengali hate speech and then train LSTM and subsequently, employ parameter sharing based transfer learning to Bengali sentiment classifiers by reusing and fine-tuning the trained weights of Hindi classifiers with both classifier being used as baseline in their study. Finally, they use BiLSTM with self attention in joint dual input learning setting where we train a single neural network on Hindi and Bengali dataset simultaneously using their respective embeddings.

Chapter 3

Preliminaries

3.1 Deep Learning

In the 1980s, Noel Entwistle and colleagues proposed the term deep learning for the first time in research about how to distinguish deep and surface learning. Artificial neural networks have become a branch of machine learning, deep learning can be divided into supervised, semi-supervised or unsupervised learning. The main concept in deep learning algorithms is the automated extraction of representations from data. Another key concept closely related to deep learning method is learning the distributed representation of data. In this case, each sample can be represented compactly, leading to a richer generalization.

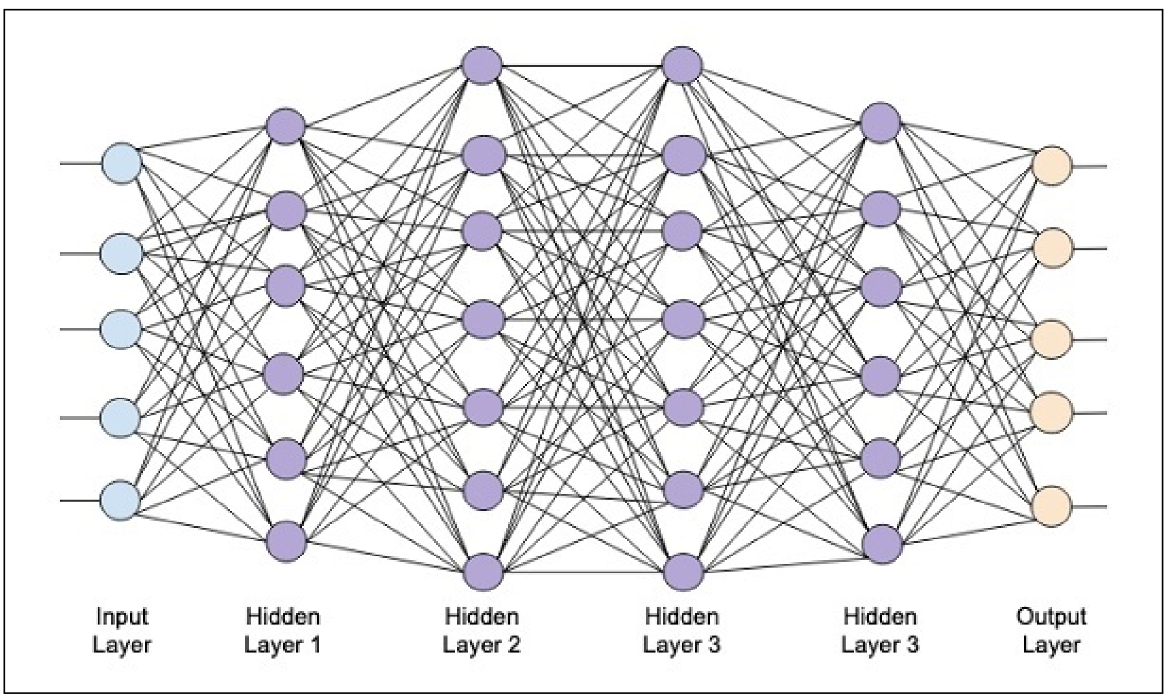

Figure 3.1: Architecture of example deep learning model.

For deep learning algorithms, it can be simply thought of as deep architectures of consecutive layers (see Figure 3.1). Each layer applies nonlinear transformation to its input and output the representation. The input layer contains your input data. The hidden layer is trying to learn different aspects of the data by minimizing an error or cost function. The output layer consists of the output data. The purpose is to learn a complex and abstract data representation hierarchically by passing data through multiple transformation layers. The input data are fed to the first layer such as token (word piece) embeddings. The output of each layer is the input of the next layer.

The basic idea in deep learning algorithms is stacking up the nonlinear transformation layers. More complex nonlinear transformations can be constructed from deeper layers. Through a deep architecture with multiple levels of representations, the data are transferred into abstract representations. In this case, deep learning algorithms can be considered as a kind of representation learning algorithm.

The final trained model can be thought of as a highly nonlinear function of the input data which can construct a final representation. The underlying explanatory factors in the data can be extracted from the nonlinear transformations by the layers of deep architecture.

The final representation (the output of the final layer) contains the useful information in the training data, constructed by the deep learning algorithm, which can be used as the features in building classifiers in a high-efficiency comparing with the high dimensional sensory data.

In this project, we used a pre-trained language representation model built in deep learning techniques to extract the information from text and transfer the text into vectors. The output of the language model called work embeddings containing the information of the input text.

Pre-trained Language Models: Due to the effectiveness of the pre-trained language model in many NLP downstream tasks, it has received much attention. Language model pre-training has been proved the efficiency for improving many natural language processing tasks such as sentiment classification (Xipeng, et al., 2020). The basic idea behind the pre-trained language model is training a word embedding layer from a large scale of style so that it has an excellent ability to extract the information from contextual text. Because it is not enough to train various neural architectures of coding context representation only from the limited supervision data of terminal tasks. Bidirectional Encoder Representations from Transformers (BERT) is a pre-trained language representation model proposed based on deep learning techniques by Google AI team in 2018 (Devlin, et al., 2018) [3]. Different with other language representation models, through jointly condition on both left and right context in all layers, BERT can generate deep bidirectional representations from the unlabeled input text. BERT has been applied in various NLP tasks such as text classification and question answering and preformed an excellent performance (Yuwen and Zhaozhuo, 2018). Because of the fine-tuning approach adopted, there is no specific architecture for downstream NLP tasks when we use BERT. As an intelligent agent, it should minimize the use of prior human knowledge in the model design and learn such knowledge from data instead of. In BERT, there are two different objectives used to train the language model rather than the frequently used objective of next-word prediction: The first is the masked language model objective, where the model needs to predict the masked tokens from their context. The other one is the next-sequence prediction objective, where the model needs to learn whether sequence B follows sequence A. Those two objectives enable the model to learn long-term dependencies better.

- **Masked Language Model Objective:** The model learns to predict the tokens masked out randomly in sequence A and sequence B.
- **Next-Sentence Prediction:** In order to enable BERT to learn long-term dependencies better, the model needs to learn if a sequence B would naturally follow the previous sequence A. So the sequence A and sequence B are from the same document so that sequence A follows sequence B.

In BERT (Devlin et al., 2019) [3], the authors use the transformer as basic components rather than recurrent or convolutional neural networks. The transformer is solely based on the self-attention mechanism. Compared with Recurrent Neural Network (RNN) or Convolutional Neural Network (CNN), the transformer has three advantages. Firstly, it can reduce the computation resource and computation speed. Secondly, the computation can be parallelized which is impossible in RNN. Otherwise, the transformer has a good performance in learning long-range dependencies. In practice, it is easy to create an excellent performance model by fine-tuning the BERT with one additional output layer for various NLP tasks such as text classification and question answering, without too much substantial task-specific architecture modifications (Devlin et al., 2019) [3].

3.2 Sentiment Analysis

Sentiment analysis (SA, also known as opinion mining) is defined as a computational task of finding people's opinions about specific entities. There are three main classification levels in sentiment analysis (Medhat and Hassan, 2014): document-level, sentence-level, and aspect-level sentiment analysis. The purpose of document level emotion analysis is to classify opinion documents as expressing positive or negative opinions or emotions. It considers the whole document as a basic unit of information. Sentence-level SA aims to classify sentiment expressed in each sentence. Actually, we can think of sentences as a short document so that there is no fundamental difference between document-level and sentence-level (Liu B., 2012). Aspect-level sentiment analysis will be discussed in this thesis.

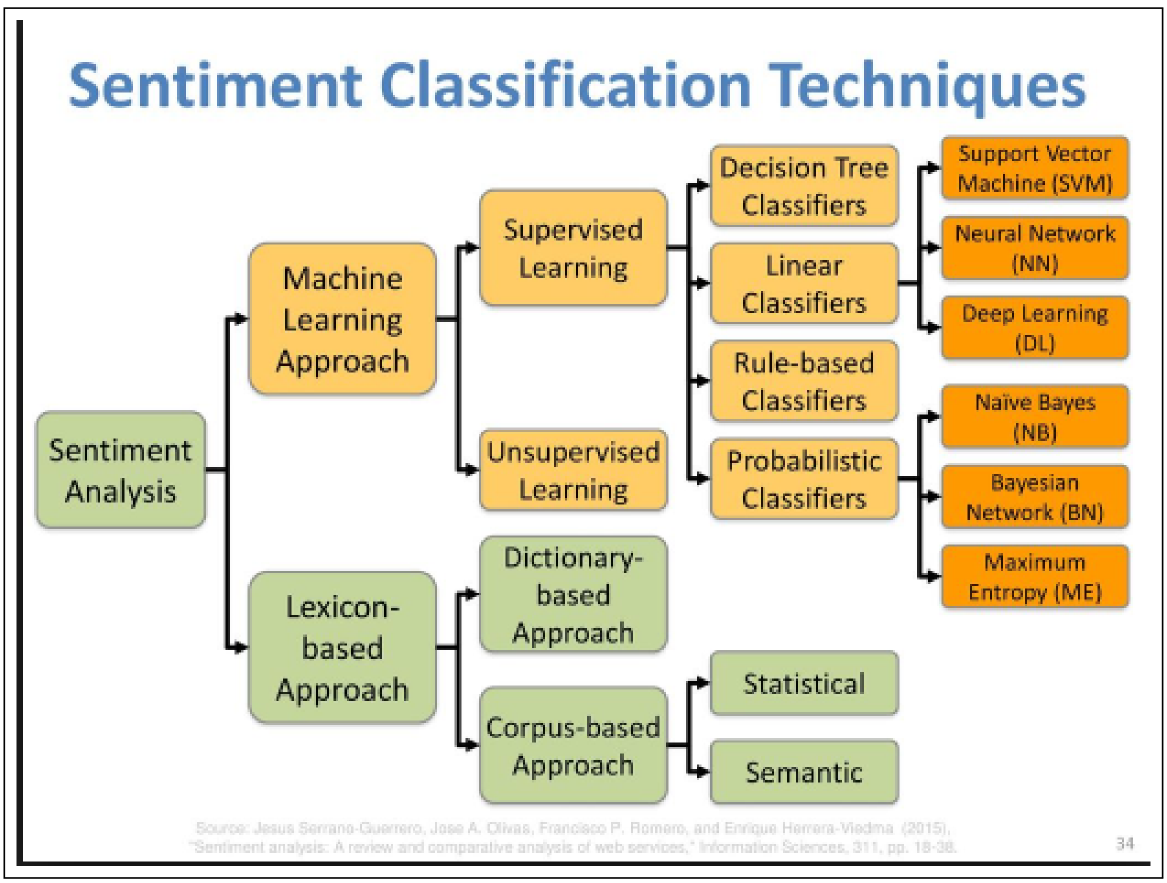

Figure 3.2: Sentiment Types

Sentiment classification techniques can be divided into two categories: The detailed algorithms are showing in Figure 3.2. Sentiment Classification techniques can be roughly divided into the hybrid approach, machine learning approach and the lexicon-based approach. The Machine Learning (ML) approach applies ML algorithms and uses linguistic features. The lexicon-based approach relies on a sentiment lexicon, which is a collection of pre-compiled and known sentiment terms. More detailed, it can be divided into the dictionary-based approach and corpus-based approach which use statistical and semantic methods to find sentiment polarity, respectively. The hybrid approach combines both approaches played a critical role in the majority of the methods which is very common with sentiment lexicons. The sentiment classification method using lexicon-based approach can be divided into the dictionary-based approach and the corpus-based approach, which depends on finding the sentiment lexicon. The dictionary-based approach begins with finding sentiment or opinion seed words and then searches the dictionary of their synonyms and antonyms. The corpus-based approach depends on a seed list of opinion words and then finds other sentiment words using statistical or semantic methods in a large corpus to help in finding sentiment words with context-specific orientations.

Machine learning approaches are the dominant approaches in the sentiment analysis task. It depends on the features of data when used to sentiment analysis. There are two approaches: unsupervised and supervised learning methods. The supervised methods make use of a large number of labeled training documents. The unsupervised methods are used when it is difficult to find these labeled training documents when they do not exist.

The Bag Of Words (BOW) model is a traditional ML approach that is frequently used. The main idea is to map feature vectors from a document and then classify by machine learning techniques. Despite the simplicity and efficiency of the BOW method, a lot of the information from the original natural language is lost because various types of features have been exploited, such as word order and syntactic structures. In general, traditional approaches such as Support Vector Machine (SVM) is based on complex manually extracted feature, which is a time-consuming and complex process. Traditional machine learning methods contain many steps and fundamental questions like complex features extraction from text data, figuring out the relevant features, and selecting a suitable classification algorithm for the tasks.

Deep learning is an increasingly popular alternative to traditional machine learning methods because of its excellent performance in Natural Language Processing (NLP) tasks such as sentiment analysis. Compared with traditional methods, more complex features can be extracted from the data automatically when using neural networks but with minimum external contribution. Figure 3.2 shows a clear difference between those different techniques: Compared with traditional methods, deep learning could help to extract features automatically rather manually. While deep learning techniques have been applied to many NLP tasks, usually those models required large datasets and high-performance computational resources for training. In this thesis, BERT deep learning is used to do sentiment analysis, which belongs to neural network, performed as in Figure 4.15.

3.3 Social Media Analytics

The Internet and mobile technologies are the main forces of the rise of social media, providing a technical platform for information dissemination, content generation and interactive communication. Social media has become a critical part of the information ecosystem.

Over the past few years, the research of social media has greatly intensified by the significant interest from different domains. Social media analytics, usually driven by specific requirements by a target application, involves developing and evaluating informatics tools and frameworks to collect, monitor, analyze, summarize, and visualize social media data. Social media analytics contains a three-stage process: "capture", "understand" and "present". Figure 3.3 presents this process with explanations.

Capture: The capture stage is helpful to identify information on social media platforms related to its activities and interests by collecting enormous amounts of relevant data from many social media sources. These data are archived and available to meet the requirements of task. Through various preprocessing steps, including data modeling, data and record linking from different sources, stemming, part-of-speech tagging, feature extraction, and other syntactic and semantic operations that support the analysis, the processed data are delivered to the understanding stage.

Understand: Usually, there is a considerable part of the noisy data that exists in the data collected from many uses and sources on the capture stage, which need to be removed before meaningful analysis. Then, many techniques from machine translation, text, natural language processing, data mining and network analysis can be involved in accessing meaning from the cleaned data. At this stage, many useful metrics and trends about users can be produced, covering users' backgrounds, interests, concerns and relationship networks. Note that the understanding stage is the heart of the entire social media analysis pro-

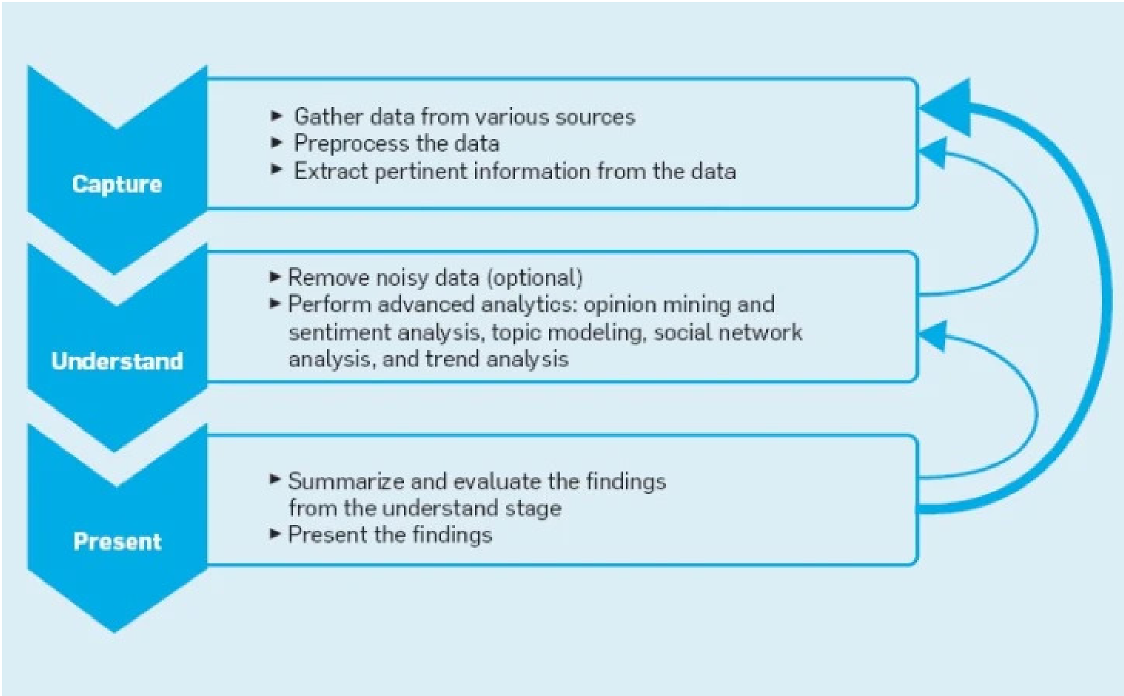

Figure 3.3: Social Media Analytics Process

cess. Its results will have a significant impact on the information and metrics in the present stage, these results will be of great help to the decision-making of businesses.

Present: As the last stage, the results from different analytics are evaluated, summarized and shown in an easy-to-understand format. Various visualization techniques can be used to present useful information. In this thesis, the main process of sentiment analysis follows the social media analytics process of Figure 3.3. The capture stage corresponds to data description and data preprocessing. The method part will mainly focus on understanding and extracting information for study goals. Lastly, the results will be presented and analyzed in Section 5, which is the present stage.

Chapter 4

Methodology

4.0.1 Problem Statement

Our objective in this work is to generate datasets for bengali sentiment analysis from existing datasets and detect Aspect-based Sentiment in a Bengali Text using BERT (Bidirectional Encoder Representations from Transformers) Deep Learning based Language Model. For each input, output is a single Sentiment label: positive, negative or neutral.

We generate total 5 multi-domain datasets from existing datasets, and the dataset we collected, usually have texts from single domain.

To address this multi-class classification problem (positive, neutral, negative), we use pre-trained BERT-base language model tranined on Bengali Language by Sagor Sarker at [12]. BERTs model architecture is a multi-layer bidirectional transformer encoder based on the original implementation described in [13] and released in the tensor2tensor library [3].

The steps that will be followed in this study is shown below diagrammatically -

4.1 Data Collection

This is the initial phase of our project. In this phase, we collect datasets from the internet shown below. Reviews/Data in these datasets are on a single topic/domain e.g. Cricket, Drama, etc. Some texts in these datasets may contain English words. Datasets are -

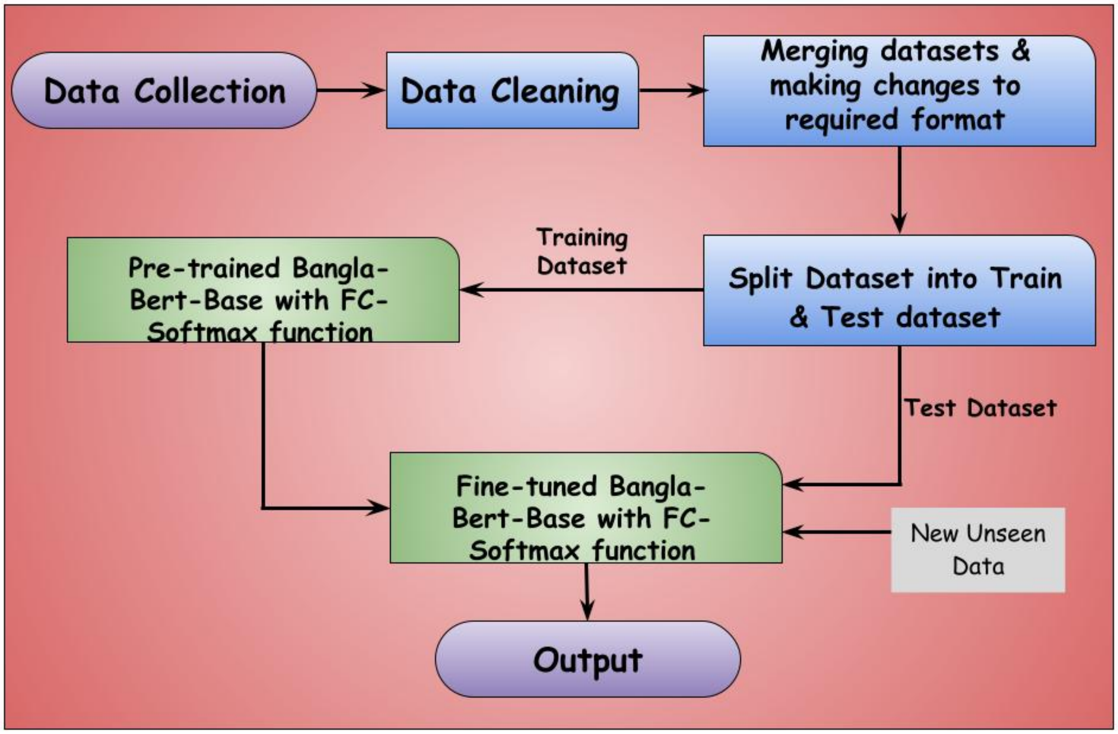

Figure 4.1: Solution Steps for Sentiment Analysis from Bengali Text

4.1.1 Bangla (Bengali) sentiment analysis classification benchmark dataset corpus

This corpus [14] contains 3307 Negative and 8500 Positive comments from Youtube about Bengali Drama stored in all_negative_3307.txt and all_positive_8500.txt files. The total no. of comments is 11807. Sentiment labels are positive and negative. Samples from datasets are shown in Figure 4.2.

index	Text	Tag
1806	কাজল আরফিন অমি মানে ভালো একটা কিছু	positive
5346	এই ধরনের ফালতু নাটক বানিয়ে এইসব কুপ্রথাকে প্রমোট করা হচ্ছে,যত্তসব ফাউল।পারলে ভাল কিছু বানান।	negative
5624	কুব অসাধারন	positive
7408	অসাধারণ ।অপূর্ব। আপনাদের নাটক অসম্ভব রকমের প্রশংশার দাবি রাখে।অনেক অনেক ধন্যবাদ।অপূর্ব ও তিসাকে ধন্যবাদ।	positive
3282	Very very nice. নাটকটির অপেক্ষায় ছিলাম ।নিশো আর তিশা মানে ভালো কিছু নতুন কিছু । জুটি দুটো সেই ।	positive
6850	অনেক বার দেখছি অনেক ভালো লাগে নাটক টা ধন্যবাদ সবাই কে	positive

Figure 4.2: Sample from Bangla (Bengali) sentiment analysis classification benchmark dataset corpus

4.1.2 KhondokerIslam/Bengali-Sentiment Dataset

This dataset [15] contains two files - train.csv & test.csv. This dataset contains texts from the cricket domain. The total no. of data is 17852. Sentiment labels are 0, 1, 2. Samples from datasets are shown in Figure 4.3.

index	Data	Sentiment
8106	পরবর্তি আকুর পেমেন্ট পরিশোধ করতে আরো ২১ দিন বাকি । আশা করি এই ২১ দিনে আরো ১০০কোটি ডলার রিজার্ভ বাড়বে । আশা করি সরকার ঢাকা চট্রগ্রাম মেট্রোরেলের প্রকল্প নিবে অচিরেই ।	1
9648	দীর্ঘকাল পর দেখলাম ' সাংবাদিকতা ' । এভাবে যদি স্বনির্বাচিত কাগুজে এমপি , মন্ত্রী ও প্রধানমন্ত্রীকে সাংবাদিকরা প্রশ্ন করত তাহলে বাপ রে বাপ বলে পালাত তারা । ফিরে আসুক এমন সাংবাদিকতা । দেশ আজ এমন ভঙ্গুর হতো না । মাশরাফি রাজনীতিবিদ নন বলেই এমন প্রশ্ন করা সম্ভব হয়েছে । রাজনীতিবিদদের বেলায়ও সাংবাদিকরা এমন সরব হোন । জয় হোক ' সাংবাদিকতা ' র ।	1
9167	এই শক্তির জোগান দিচ্ছে রিয়াল আর মিডিয়াতে বিভিন্ন যুক্তি দিয়ে আলোচনা জিইয়ে রাখছে মার্কা ।	0
6234	২০০৪ সালের ২১ আগস্টে গ্রেনেড হামলার ঘটনায় তাঁকে চেয়ারম্যান করে এক সদস্যের বিচার বিভাগীয় তদন্ত কমিশন গঠন করেছিল ওই সময়ের বিএনপি - জামায়াত জোট সরকার । বিচারপতি জয়নুল আবেদীন এ ঘটনায় বহুল আলোচিত ১৬২ পৃষ্ঠার একটি প্রতিবেদন জমা দেন । প্রতিবেদনে বলা হয় , গ্রেনেড হামলার সঙ্গে একটি শক্তিশালী বিদেশি গোয়েন্দা সংস্থা জড়িত । কিন্তু সুনির্দিষ্টভাবে কোনো গোয়েন্দা সংস্থার নাম বলা হয়নি । পরে জজ মিয়া নাটকসহ বিভিন্ন কারণে বিতর্কিত ওই প্রতিবেদনটির সত্যতা প্রমাণিত হয়নি ।	0
12215	মাঝখান থেকে লুকমানের ঘুম কামাই	0
450	ব্যক্তিগত বিনিয়োগ আর উদ্দোক্তা সৃষ্টিতে সরকার বরাবরের মতই বিফল	2

Figure 4.3: Sample from KhondokerIslam/Bengali-Sentiment Dataset

4.1.3 eftekhar-hossain/Bengali-Restaurant-Reviews

This dataset [16] is made from Bengali Restaurant Reviews and contains a total of 1431 reviews. Sentiment labels are positive and negative. Samples from datasets are shown in Figure 4.4.

index	Reviews	Sentiment
885	এ সব তাদের খাদ্য কোন ভালো লাগে নি। মেনু দিয়ে পৃষ্ঠাগুলির পূর্ণ কিন্তু মান সামান্য । আবার যাওয়ার কথা মনে করবেন না।"	negative
735	একমাত্র সমস্যা হল ম্যানেজার সম্পূর্ণ অসম্পূর্ণ।	negative
78	প্রায়ই এখানে আসা হয় নিয়মিত গ্রাহকের মত। খাবারের মান এবং পরিবেশ দুইটাই ভাল। তাদের পিজা এবং সেট মেনুর দুইটাই জোস ..	positive
249	তাদের খাবার খুব ভাল ছিল ... কিন্তু তারা সেই সুস্বাদু খাবার পরিবেশন করার জন্য অনেক সময় নেয় .... সামগ্রিকভাবে এটি Hangout এর জন্য একটি ভাল জায়গা ..	positive
1247	কি একটি বিস্ময়কর restaurent !! এই খাদ্য স্বাদ ঘৃণ্য হয়, আমি তার ফুটন্ত উপর এবং মসলাযুক্ত মত আমার সমগ্র খাদ্য স্বাদ life.I অর্ডার পাস্তা একটি খারাপ অভিজ্ঞতা ছিল, লবণ etc.chicken ভাজা হয়ভয়ানক .আমার মূল্যবান টাকা নষ্ট.The মূল্য খুব বেশী হয় কিন্তু খাবার খুব খারাপ .	negative
1122	তাদের খাদ্য স্বাদ poor.সার্ভিস এছাড়াও শূন্য হয়।	negative

Figure 4.4: Sample from eftekhar-hossain/Bengali-Restaurant-Reviews dataset

4.1.4 eftekhar-hossain/Bengali-Book-Reviews

This dataset [17] is made from Bengali Book Reviews and contains a total of 1444 reviews. Sentiment labels are 0 and 1. Samples from datasets are shown in Figure 4.5.

index	Review	Sentiment
1417	ফাটাফাটি একটা বই	1
728	মার্জিত রুচি, বুদ্ধির দীপ্তি, অপূর্ব বাকচাতুর্য ও স্মিথ রসিকতায় পরিপূর্ণ ‘পূর্বের চোখ পশ্চিমের মন’ পাঠক হৃদয়ে অমর হয়ে থাকবে ।	1
1159	ভালো অনুবাদ, খুব ভালো	1
767	আলহামদুলিল্লাহ অনেক ভালো বই।	1
329	এই বইটি পড়ার সময় আমার মানসিক অবস্থা লোপ পাওয়ার উপক্রম হয়েছিলো। তাই সবাই নিজ দায়িত্বে বইটি পড়বেন।	0
133	Pure trash. এইটার নাম দেয়া উচিত ছিল 'আমার আছে ঢং'। নায়ক-নায়িকা থেকে শুরু করে সব কয়টা চরিত্র ঢং-ঢাংয়ের উপরেই থাকে। এই বইগুলির জন্য অন্তত গুডরিডসে নেগেটিভ রেটিং দেয়ার ব্যবস্থা থাকা উচিত ছিল।	0

Figure 4.5: Samples from eftekhar-hossain/Bengali-Book-Reviews dataset

4.1.5 AtikRahman/Bangla-ABSA-Datasets/Cricket

This dataset [18] contains a total of 2979 comments on Cricket. Sentiment labels are positive, negative, and neutral. Samples from datasets are shown in Figure 4.6.

index	Source	Date	Text	Category	Polarity
2651	Prothom Alo	2018-10-02 00:00:00	আন্তর্জাতীক টিমে আসলে তাদের টাউজার ফেটে যাই.....হা হা হা??	team	negative
1288	Prothom Alo	2018-11-02 00:00:00	এরকম ভাল সিদ্ধান্তের জন্য তামিম-মাশরাফি ও বিপিএল গভর্নিং বডিকে সাধুবাদ জানাই	team management	positive
2586	Prothom Alo	2018-10-02 00:00:00	প্লেয়ার দের নামে বাজে কমেন্ট করিস, তোর বোদ্দগুষ্টি মাশরাফির পায়ের কাছে যেতে পারবে	other	negative
74	BBC Bangla	2018-08-02 00:00:00	সালা সিলেক্টারদের জুতা মারা উচিত। ভালো খেলোয়ারদের সুযোগ দেয়না	team management	negative
680	Prothom Alo	2018-08-02 00:00:00	জ শ্রীলঙ্কান স্পিনার হেরাত ৪০/৪২ বছর বয়সে ক্রিকেট খেলছে!!	bowling	neutral
622	Prothom Alo	2018-04-02 00:00:00	মুশফিকের আমলে বিকল্প প্লেয়ার কম ছিল	batting	neutral

Figure 4.6: Samples from AtikRahman/Bangla-ABSA-Datasets/Cricket dataset

4.1.6 AtikRahman/Bangla-ABSA-Datasets/Restaurant

This dataset [19] contains a total of 2059 reviews from the Restaurant domain. Sentiment labels are positive, neutral, negative, and conflict. Samples from datasets are shown in Figure 4.7.

index	SL	Text	Category	Polarity
871	874	তাই যদি আপনি Montparnasse একটি সুন্দর, উপভোগ্য খাবার চান, প্রাক থিয়েটার প্রিক্স-ফিক্সের জন্য প্রারম্ভিক যান।	price	positive
1031	1034	এটা ব্যাংক ভাঙ্গবে না এবং আমিও খাবারের জন্য ফিরে আসব না।	food	negative
1902	1956	বৃহত্তম সুশি এবং সুস্বাদু এবং সুস্বাদু টুকরো টুকরা,	food	positive
1438	1442	বাগানের পিছনের বসার জায়গা খুব শানতির ছিল যেখানে আপনি তাদের নিজেশব বাগান দেখতে পারবেন	anecdotes/miscellaneous	neutral
1409	1413	কিন্তু আপনি বসেছিলেন ওয়েটাররা ভালই ছিল তারা মেনুতে সবকিছুর ব্যাখ্যা আপনি পাবেম, এবং খাদ্য মূল্য থেকে সেবা জন্য সত্যিই সস্তা।	price	positive
1513	1517	আমি সেখানে তিনবার গিয়েছি এবং সর্বদা চমৎকার অভিজ্ঞতা হয়েছে।	anecdotes/miscellaneous	positive

Figure 4.7: Samples from AtikRahman/Bangla-ABSA-Datasets/Restaurant dataset

4.1.7 Data Set For Sentiment Analysis On Bengali News Comments

This dataset [20] contains a total of 13802 reviews and Sentiment labels are কিছুটা নেতিবাচক, নিশ্চিত নেতিবাচক, নিরপেক্ষ, কিছুটা ইতিবাচক, নিশ্চিত ইতিবাচক. Samples from datasets are shown in Figure4.8.

index	data	title_x	title_y	title	value	tag
9625	সব কিছু নেগেটিভ দেখেন কেন? এটার মানে তার পারফরমেন্স, ফিটনেস বা অন্য কিছুও হতে পারে। অতএব সহজ চিন্তা করুন, কুটিলতা পরিহার করুন।	-1	2	2	3	নিশ্চিত ইতিবাচক
5068	বিএনপি দলটির ভবিষ্যত আসলে কি? তারা জনগণের জন্য রাজনীতি করেনা, তারা রাজনীতি করে নিজের লাভের আশায়। ভুল, ভন্ডামী কাণ্ড ও মিথ্যাচারের আবর্ত থেকে দলটির বেরিয়ে আসার কোন লক্ষনই দেখা যাচ্ছে না! বিম্পি'র রাজনীতিতে ভাল কি আছে?	-1	1	-2	-2	নিশ্চিত নেতিবাচক
9616	২০০৭ বিশ্বকাপের সর্বোচ্চ উইকেট শিকারী ম্যাকগ্রা স্ত্রীর স্তন ক্যান্সার হওয়া সত্ত্বেও সব মনোযোগ বিশ্বকাপে রেখেছিলেন দেশকে বিশেষ কিছু দেয়ার জন্য,রাজনীতিতে যোগ দেননি।তাই আপনি ম্যাকগ্রা হবেন কিভাবে?	0	-1	-2	-3	নিশ্চিত নেতিবাচক
8889	এটা নিয়ে নিউজ করার কি আছে	-1	1	0	0	নিরপেক্ষ
3442	বাবা-মা উভয় আমাদের জন্য জান্নাত অথবা জাহান্নাম । তাই আমাদের উচিত তাঁদের সাথে সদ্ব্যবহার করা ।	2	2	2	6	নিশ্চিত ইতিবাচক
12378	নাসিম বিনাভোট এ জিয়া এসেছিল । ক্ষমতা বন্দুকের নলে । আঃলীগ ভোটের মধ্যে আসেছে । পিছনের দরজা দিয়ে নয়	-1	2	0	1	কিছুটা ইতিবাচক

Figure 4.8: Samples from Sentiment Analysis On Bengali News Comments

4.2 Data Cleaning

Data is the most important thing in any Deep Learning based work. Without proper data, it is very much impossible to create deep learning based trained model which produces correct outputs. Data collected from different online platforms contains emojies, extra irrelevent data or missing some part of data.

Data cleaning is the process of identifying the incorrect, incomplete, inaccurate, irrelevant or missing part of the data and then modifying, replacing or deleting them according to the necessity.

For removing emojies from data, we use demoji python library. Also we can use regex or re python library for removing unwanted segments of data or incomplete data.

4.3 Generating Multi-domain datasets

To create our dataset, we combine these datasets based on No. of classes of output (e.g. pos, neg, and/or neutral) and generate 5 datasets. How we generate the datasets are given below -

4.3.1 Multi-Domain Dataset 1

For generating Multi-Domain dataset-1, we rename the Tag column of Bangla (Bengali) sentiment analysis classification benchmark dataset corpus to Sentiment. Removed rows of KhondokerIslam/Bengali_Sentiment Dataset having sentiment value of 1 and changed sentiment labels 0 to negative and 1 to positive. Renamed the Data column to Text. Then, we merged these two datasets. Resultant Merged dataset 1 has sentiment values of positive, and negative. Samples from datasets are shown in Figure 4.2.

index	Text	Sentiment
21760	এই প্রথম ভালো এক জন পুলিশ দেখলাম	positive
6698	কলা বাবা কলা দিবে কলা খাইয়া যা কামর দিয়া মনে মনে যা খুশি তা চা গানটা অনেক সুন্দর লাগছে ।	positive
16671	দু দিন পর পর টাকা খাওয়ার প্রকল্প	negative
1305	দারুন চোখ দিয়ে জল বেরিয়ে গেলদারুন অভিনয় করেছেন ভাই	positive
14476	মোদির আস্কারায় মৌলবাদীরা ভারতে অপ্রতিরোধ্য হয়ে উঠছে ।	negative
9328	খুব সুন্দর এক্টা নাটক। বিশেষ করে নিশো ভাইকে ধন্নবাদ এতো সুন্দর এক্টা মেসেজ দেবার জন্নে যেটা হচ্ছে নামাজ!!! আর এই নাটকের সব থেকে বড় বিউটিই হলো নামাজের খবরটা।:) ওমি ভাই আল্লাহ আপ্নাকে ভালো রাখুক সব সময়।:)দোয়া করি।আরো ভালো ভালো মেসেজ দিবেন ভাই।দুনিয়াটা অস্থায়ী ভাই।:)	positive

Figure 4.9: Samples from Multi-Domain Dataset 1

4.3.2 Multi-Domain Dataset 2

For generating Multi-Domain dataset-2 - From KhondokerIslam/ Bengali_Sentiment Dataset changed the column name Data to Text and changed sentiment values 0, 1, 2 to negative, neutral, and positive, respectively. From AtikRahman/ Bangla _ABSA _Datasets/ Cricket removed the Source, Date, Category columns and renamed the Polarity column to Sentiment. From AtikRahman/ Bangla _ABSA _Datasets/ Restaurent removed the SL, Category columns and renamed the Polarity column to Sentiment. From Data Set For Sentiment Analysis on Bengali News Comments removed the title_x, title_y, title, value columns and change column names Data, Tag to Text, Sentiment, respectively. After that we change কিছুটা নেতিবাচক, নিশ্চিত নেতিবাচক to 'negative', 'নিরপেক্ষ' to 'neutral', and কিছুটা ইতিবাচক, নিশ্চিত ইতিবাচক to 'positive'. Then we merge these datasets. Samples from resultant dataset is shown in Figure 4.10

index	Text	Sentiment
26699	এই খবরে তাও যদি লাফারজের শেয়ার এর যদি কিছুটা বাড়ে ।	positive
31995	পরিসংখ্যান, রাঙ্কিং, বর্তমান, প্রাক্তন খেলোয়াড়, ধারাভাষ্যকার, সমর্থক, দর্শক সবাই তাই বলছে। বর্তমান কোনো খেলোয়াড়ের দম নেই ওরকম খেলা এটাই সত্যি।	negative
35861	মুক্তির জন্য -ধানের শীষে ভোট দিন ।	neutral
17189	অতিরিক্ত পুলিশ সুপার প্রমাণ ছাড়াই বলে দিলো , দিল মোহাম্মদকে বিএনপি - জামায়াতের লোকজন কুপিয়ে হত্যা করেছে বলে শুনেছি	negative
30770	বিএনপি জামাতিদের যে মিথ্যা আর গুজবই একমাত্র সম্বল সেটা তাদের হাবভাবেই বোঝা যায়। ভাই ইসলামি ব্যাংক কবে সরকারি পরিচালনায় গেল? এই ব্যাংকের মালিকানা পরিবর্তন হয়েছে। আগের কিছু মালিক তাদের সেয়ার বিক্রি করেছে, অন্যরা তা কিনে নিয়ে নতুন মালিক হয়েছে। তারমানে এই নয় সরকার এটার পরিচালনা করছে।	negative
2385	প্রিয় প্রথম আলো সম্পাদক	neutral

Figure 4.10: Samples from Multi-Domain Dataset 2

4.3.3 Multi-Domain Dataset 3

For Multi-Domain dataset 3 - From Data Set For Sentiment Analysis on Bengali News Comments removed the title_x, title_y, title, value columns and changed column names Data, Tag to Text, Sentiment, respectively. It is then considered as Merged Dataset 3, though a single dataset is used to generate it. Samples from the resultant dataset is shown in Figure 4.11.

4.3.4 Multi-Domain Dataset 4

For generating Multi-Domain Dataset 4, we follow the steps used to generate Multi-Domain dataset 2 except, before merging the datasets, we remove sen-

index	Text	Sentiment
3743	ইঞ্জিনিয়ারিং ঢাকা সিটি কর্পোরেশন এর মত করে কিন্তু আমরা ঢাকা সিটির বেলায় কী পেলাম? ইলেকশন নাকি সিলেকশন?	নিশ্চিত নেতিবাচক
1646	সুপ্রভাত বাংলাদেশ। সৌদি বিনিয়োগ প্রসংগে দুই মাস প্রস্তুতিমূলক কার্যক্রম সফল হউক আশা করছি।	কিছুটা নেতিবাচক
3912	কৌশলী বিএনপির কাছে ধরাশায়ী আওয়ামী লীগ। নির্বাচন হবে তো?	নিশ্চিত নেতিবাচক
3636	মনে হলো, আমি যেন ওখানেই ছিলাম আমার দেশের ভাইদের সাথে আহা	নিশ্চিত ইতিবাচক
2513	সরকারি সিলেট ইঞ্জিনিয়ারিং কলেজ	নিরপেক্ষ
795	মূল কারণ হচ্ছে সৌদি আরবে অনেক আইন করে ফেলেছে। আগের মত আর নাই।	কিছুটা নেতিবাচক

Figure 4.11: Samples from Multi-Domain Dataset 3

timent labels from dataset 7. Samples from the resultant dataset is shown in Figure 4.12.

index	Text	
27506	জনগন উন্নয়নের সুফল ভোগ করছে বলেই কেউ আর বিএনপির কথায় এখন নাচে না। এটাকেই বলে প্রতিফল।	positive
6202	খুবই দুঃখ পেলাম । আমারা যারা ৮০ দশকের শেষের দিকে এবং ৯০ এর দশকের শুরুতে বড় হয়েছি তাদের চোখে কায়সার হামিদ , সাব্বির , মোনেম মুন্না , কানন , আলফাজ , আসলাম , মহসীন প্রমুখ স্বপ্নের মানুষ ছিলেন । খেলার ২ / ৩ ঘন্টা আগেই স্টডিয়ামে চলে যেতাম শুধু মাত্র খেলোয়াড়দের বাস থেকে নামা দেখতে । একবার মোহামেডান কোন একটা বিদেশি দলের সাথে খেলছিলো । খেলায় আঘাত পেয়ে কায়সার হামিদের মাথা ফেটে যায় । উনি খেলা থামান নি , সেই অবস্থাতেই মাথায় ব্যান্ডেজ বেঁধে খেলছিলেন । মোহামেডানের সাদা জার্সি তার রক্তে লাল হয়ে গিয়েছিলো । ভাই , আপনারা জানেন না আপনাদের আমরা কোন সম্মানের আসনে তুলে রেখেছি । আপনাদের এমন অধঃপতন সত্যি আমাদের কষ্ট দেয় ।	negative
30888	বাংলাদেশের সবচেয়ে বড় অপশক্তি আ লীগ। আপনাদেরই ধ্বংস করে দেয়া দরকার দ্রুত।	negative
126	অমিয়ম করার শাস্তি অপসারন	negative
4969	জ্বিনা , রাজিব , ফয়সসালরা অলরেডি এসব সন্ত্রাসিদের পক্ষে সাফাইগীতি গাওয়া শুরু করে দিয়েছে ।	negative
12562	একটা নতুন প্রযুক্তি দেখাইল বাংলাদেশে থাকতেই হইব এমন তো কোন কথা নাই	neutral

Figure 4.12: Samples from Multi-Domain Dataset 4

4.3.5 Multi-Domain Dataset 5

For generating Multi-Domain dataset 5, we use Merged dataset 3 and remove rows having কিছুটা নেতিবাচক, কিছুটা ইতিবাচক sentiment values. Samples from the resultant dataset is shown in Figure 4.13.

4.4 Split dataset for Training & Testing

The train-test split is a technique for evaluating the performance of a machine/deep learning algorithms on new data. The procedure involves taking a dataset and

index	Text	Sentiment
7608	ধড়পাকড়ের ভয়ে নিরব বিপ্লব হবে। নৌকার স্টীকার হাতে নিয়ে নিরবে যাকে খুশি তাকে ভোট দিবেন। সেটা নৌকা হলে নৌকা, অন্যটা হলে অন্যটা। অধিক উপস্থিতির হারই এখন একমাত্র হাতিয়ার।	নিশ্চিত ইতিবাচক
37	জনগনের সরকার জনবান্ধব বাজেট দিবে এটাই স্বাভাবিক, কারন জনগন কাজ ও উন্নয়ন দেখে ভোট দিবে।	নিশ্চিত ইতিবাচক
3961	সেই সাথে আওয়ামী-লিগেরও অনেকেও চায়।	নিরপেক্ষ
5756	অত্যন্ত চমৎকার উদ্যোগ। আমি নিজেও কয়েক বার ভুগেছি একই সমস্যায়। নিজে নিয়মিত রক্ত দেওয়া সত্ত্বেও নিজের প্রয়োজনে রক্ত খুজে পেতে খুব বেগ পেতে হয়। এ্যাপটি ব্যবহার করতে পারলে হয়ত তা বহু গুন সহজ হয়ে যাবে।কিন্তু এই এ্যাপটি এখনো এ্যাপল স্টোরে দেওয়া হয়নি। যার কারনে ব্যবহার করতে পারছিনা। বিষয়টি উদ্ভাবক বন্ধুদের বিবেচনায় রাখতে অনুরোধ করছি। সবিশেষ উদ্যোক্তা বন্ধুদের কে কৃতজ্ঞতা এবং ধন্যবাদ জানাচ্ছি।	নিশ্চিত ইতিবাচক
8100	ডাকসুর দাবির মাধ্যমে নিজের জীবনকে হুমকির মধ্যে ফেলে দিয়েছেন।	নিরপেক্ষ
5769	স্যামসাংয়ের সার্ভিস চরম বাজে । কাস্টমার বিহেভিয়ারও খারাপ ।	নিশ্চিত নেতিবাচক

Figure 4.13: Samples from Multi-Domain Dataset 5

dividing it into two subsets. The first subset is used to fit the model and is referred to as the training dataset. The second subset is not used to train the model; instead, the input element of the dataset is provided to the model, then predictions are made and compared to the expected values. This second dataset is referred to as the test dataset. The train-test split is not appropriate when the available dataset is small. The reason is that when the dataset is split into train and test sets, there will not be enough data in the training dataset for the model to learn an effective mapping of inputs to outputs. There will also not be enoughdata in the test set to effectively evaluate the model performance. In our case, we will use 70% 80% of the dataset for training and remaining for testing.

4.5 BERT(Bidirectional Encoder Representation from Transformer)

BERT [3] makes use of Transformer [13], an attention mechanism that learns contextual relations between words (or sub-words) in a text. It uses multi-layer bidirectional transformer encoders for language representations. Based on the depth of the model architecture, two types of BERT models are introduced namely Bert-base and Bert-large. The Bert-base model uses 12 layers of transformers block with a hidden size of 768 and number of self-attention heads as 12 and has around 110M trainable parameters. On the other hand, Bert-large model uses 24 layers of transformers block with a hidden size of 1024 and number of self-attention heads as 16 and has around 340M trainable parameters.

4.5.1 Input/Output Representation:

To make BERT handle a variety of down-stream tasks, input representation is able to unambiguously represent both a single sentence and a pair of sentences (e.g., <Question, Answer>) in one token sequence. A sentence can be an arbitrary span of contiguous text, rather than an actual linguistic sentence. A sequence refers to the input token sequence to BERT, which may be a single sentence or two sentences packed together. BERT uses WordPiece embeddings in (Wu et al., 2016) [21] with a 30,000 token vocabulary. The first token of every sequence is always a special classification token ([CLS]). The final hidden state corresponding to this token is used as the aggregate sequence representation for classification tasks. Sentence pairs are packed together into a single sequence. Sentences are differentiated in two ways. First, separate them with a special token ([SEP]). Second, add a learned embedding to every token indicating whether it belongs to sentence A or sentence B. For a given token, its input representation is constructed by summing the corresponding token, segment, and position embeddings. A visualization of this construction can be seen in the Figure 4.14 for Sentence - “এখন বর্ষাকাল। সকাল থেকে বৃষ্টি হচ্ছে।”.

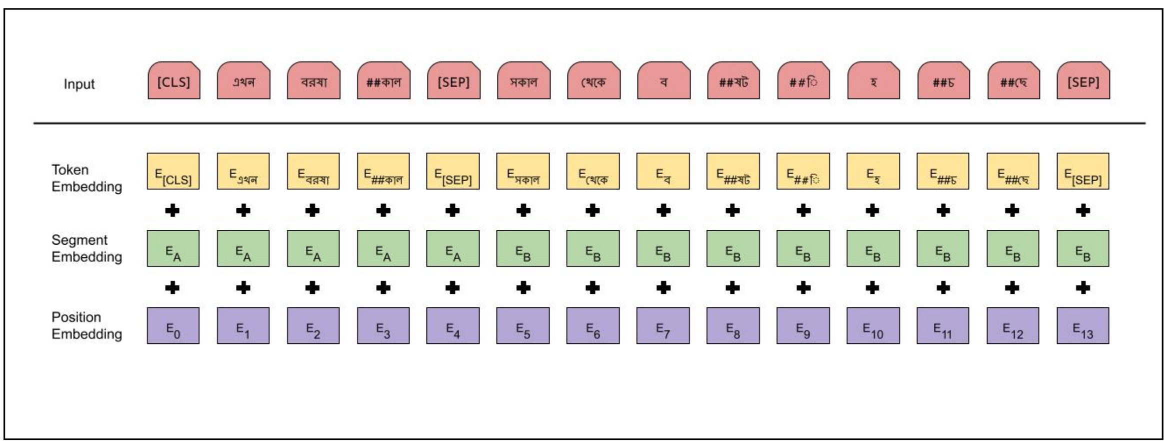

Figure 4.14: BERT input/output representation

4.5.2 Pre-training:

The methods used in the pre-training of BERT aims to generate generalized features for bidirectional language representation. Furthermore, BERT shows that this approach reduces the need for feature-engineered task-specific architectures by transfer the learned word representation and fine-tune it to other tasks. Generalized methods are crucial to achieve a model that can be utilized in transfer learning and BERT contributes with two pre-training tasks named Masked Language Model, which is a more complex version of the normal language model described, and Next Sentence Prediciton. These two tasks will be further explained in Below Sections.

Masked Langauge Model: BERT uses a mask token [MASK] to pre-train deep bidirectional representation for the language model. But the difference from the normal conditional language model that trains left-to-right or right-to-left prediction of words, where the pre-dicted word has positioned the end or start of the text sequence, BERT masks a random word in the sequence. The other reason for using a mask token to pre-train is because the standard conditional language model is only able to explicitly train left-to-right or right-to-left due to the words can have the masked word, from left-to-right, unmasked in the right-to-left, in a multilayered context and thus know what the masked word should be. The original BERT paper masked a word with a probability of 15% which was distributed as:

- 10% were replaced with a random token.
- 10% were left intact.
- 80% were replaced with the [MASK] token.

The reason for this is because of the conflict that otherwise would arise if the pre-training only made the model predict the mask tokens while the fine-tuning task would not contain any mask tokens. The model would then try to find mask token to predict but not find any in the fine-tuned task, which would result in bad perfor-mance. The pre-training would only make the model learn to extract the features from the mask token, which would not be much due to the masking only have a probability of 15% and thus make it converge slower. However, the mask token probability, used in the paper, showed that the language model learned to extract contextual word features instead.

Next Sentence Prediction: To understand the relationship between two text sentences, BERT has been pre- trained to predict whether or not there exists a relation between two sentences. Each of these sentences, Sentence A sad Sentence B, have their own embedding, in which we call embedding A and embedding B. An example given from the BERT paper was the following:

sentence A : [CLS] the man went to the store . [SEP] **sentence B :** he bought a gallon of milk . [SEP] **Label :** IsNextSentence

sentence A : [CLS] the man went to the store [SEP] **sentence B :** penguins are flightless [SEP] **Label :** IsNotNextSentence

During training, sentence B is the follow up of sentence A in half of the time to be used to predict IsNextSentence label. On the other half of the time, a random sentence is chosen for sentence B to predict IsNotNextSentence label.

4.5.3 Fine-tuning:

In this step, bert model is fed with the Multi-Domain train datasets. In this step, we add Fully-Connected Softmax layer at the end of Bert for classification type of problems like this. Then we evaluate the performance of Bert using suitable performance metrics described in next chapter. In our project, we use pretrained bangla-bert-base model by Sagor Sarker [12]. Overview of our model architecture is given in Figure 4.15 -

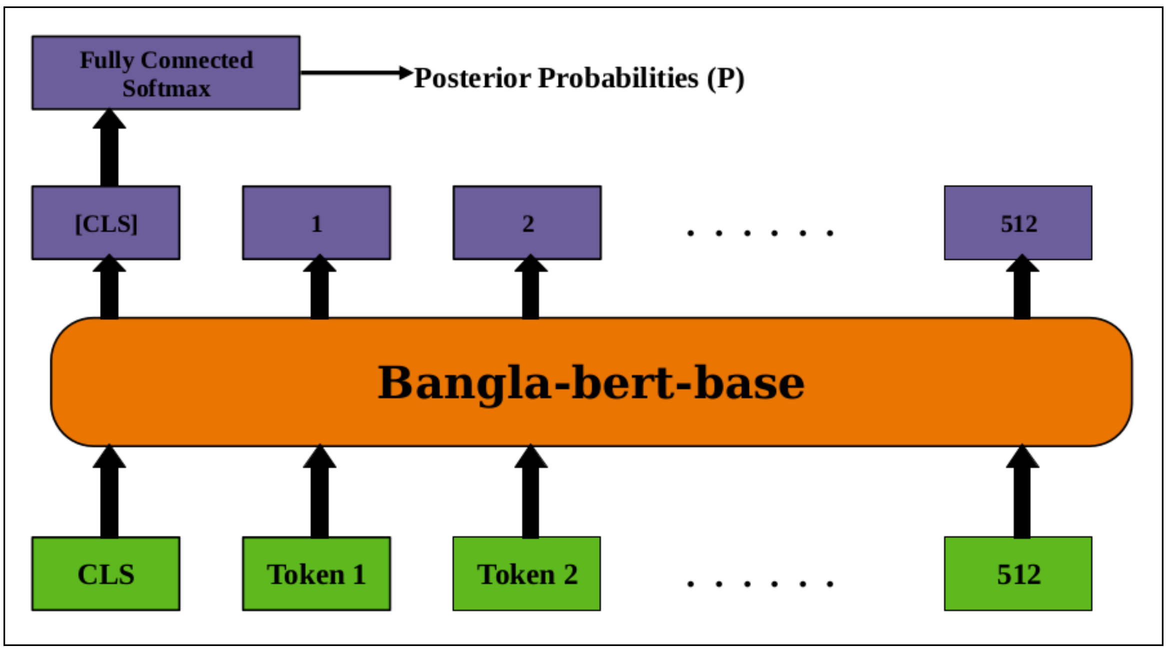

Figure 4.15: An overview of the deep learning model architecture for sentiment analysis

Chapter 5

Experiments & Results

5.1 Evaluation Metrics

In this section, the evaluation of learning models has been performed using Multi-Domain datasets. And the performance of learning models is assessed through the following metrics:

Accuracy: Accuracy is defined as the ratio between the number of correct predictions made by the model and the number of rows in the dataset. In equation 1 Tp Tn represents true positive and negative whereas Fp and fn show the false positive and negative respectively as shown -

$$Accuracy = (Tp + Tn)/(Tp + Tn + Fp + Fn)$$

Precision: Precision is defined as the ratio between the correct predictions and the total predictions yielded by the system. It can be calculated using the following formula as shown -

$$Precision = Tp/(Tp + Fp)$$

Recall: The recall is the ratio between the correct predictions made by the model and the total number of true and false sentiment labels. Recall is computed using below formula -

$$Recall = Tp/(Tp + Fn)$$

F-Measure: F-Measure is another commonly used metric in the multi-class classification task. It is defined as the geometric mean of precision and recall. F-measure is an effective metric of measuring the performance of the model as mentioned -

$$F - score = (2 * Recall * Accuracy)/(Recall + Accuracy)$$

5.2 Experiment with Multi-Domain Dataset-1

Dataset: In this experiment, we use Multi-domain dataset 1 which contains text from Bengali Drama and Cricket domains.

Model Accuracy: Accuracy of this fine-tuned model on test dataset is 0.82.

Classification Report:

	precision	recall	f1-score	support
positive	0.79	0.83	0.81	574
negative	0.85	0.82	0.83	673
accuracy			0.82	1247
macro avg	0.82	0.82	0.82	1247
weighted avg	0.82	0.82	0.82	1247

Table 5.1: Table for classification report for Multi-Domain Dataset-1

5.3 Experiment with Multi-Domain Dataset-2

Dataset: In this experiment, we use Multi-domain dataset 2 which contains text from Cricket, Restaurent, News domains.

Model Accuracy: Accuracy of this fine-tuned model on test dataset is 0.86.

Classification Report:

	precision	recall	f1-score	support
negative	0.88	0.90	0.89	870
neutral	0.82	0.77	0.80	431
positive	0.85	0.85	0.85	528
accuracy			0.86	1829
macro avg	0.85	0.84	0.84	1829
weighted avg	0.85	0.86	0.85	1829

Table 5.2: Table for classification report for Multi-Domain Dataset-2

5.4 Experiment with Multi-Domain Dataset-3

Dataset: In this experiment, we use Multi-domain dataset 3 which contains text from News domain.

Model Accuracy: Accuracy of this fine-tuned model on test dataset is 0.31.

Classification Report:

	precision	recall	f1-score	support
কিছুটা নেতিবাচক	0.37	0.39	0.89	193
নিশ্চিত নেতিবাচক	0.30	0.24	0.80	168
নিরপেক্ষ	0.30	0.30	0.85	146
কিছুটা ইতিবাচক	0.15	0.17	0.16	82
নিশ্চিত ইতিবাচক	0.37	0.38	0.38	102
accuracy			0.31	691
macro avg	0.30	0.30	0.30	691
weighted avg	0.31	0.31	0.31	691

Table 5.3: Table for classification report for Multi-Domain Dataset-3

5.5 Experiment with Multi-Domain Dataset-4

Dataset: In this experiment, we use Multi-domain dataset 4 which contains text from Cricket, Restaurent, News domains.

Model Accuracy: Accuracy of this fine-tuned model on test dataset is 0.79.

Classification Report:

	precision	recall	f1-score	support
negative	0.85	0.80	0.83	743
neutral	0.74	0.77	0.76	424
positive	0.75	0.81	0.78	430
accuracy			0.79	1597
macro avg	0.78	0.79	0.79	1597
weighted avg	0.80	0.79	0.79	1597

Table 5.4: Table for classification report for Multi-Domain Dataset-4

5.6 Experiment with Multi-Domain Dataset-5

Dataset: In this experiment, we use Multi-domain dataset 5 which contains text from News domain.

Model Accuracy: Accuracy of this fine-tuned model on test dataset is 0.52.

Classification Report:

	precision	recall	f1-score	support
নিশ্চিত নেতিবাচক	0.60	0.61	0.60	193
নিরপেক্ষ	0.45	0.44	0.45	153
নিশ্চিত ইতিবাচক	0.46	0.47	0.46	112
accuracy			0.52	458
macro avg	0.50	0.51	0.0	458
weighted avg	0.52	0.52	0.52	458

Table 5.5: Table for classification report for Multi-Domain Dataset-5

Chapter 6

Conclusion

In this project, we fine-tuned same BERT model for 5 different multi-domain datasets and compare their results. We found out that the domains which uses similar terms to describe similar expression, BERT gives more accuracy.

In future, we will collect more data from different domains and increase Multi-domain dataset size and No. of domains. And we also like to test these datasets on other Langauage models such as DistilBERT, XLNet, etc.

Bibliography

1. M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer, "Deep contextualized word representations," *CoRR*, vol. abs/1802.05365, 2018. Accessed: 2022-07-12.

2. J. Howard and S. Ruder, "Fine-tuned language models for text classification," *CoRR*, vol. abs/1801.06146, 2018. Accessed: 2022-07-12.

3. J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018. Accessed: 2022-07-10.

4. M. Hoq, P. Haque, and M. N. Uddin, "Sentiment analysis of bangla language using deep learning approaches," in *International Conference on Computing Science, Communication and Security*, pp. 140–151, Springer, 2021. Accessed: 2022-07-10.

5. A. Ahmed and M. A. Yousuf, "Sentiment analysis on bangla text using long short-term memory (lstm) recurrent neural network," in *Proceedings of International Conference on Trends in Computational and Cognitive Engineering*, pp. 181–192, Springer, 2021. Accessed: 2022-07-10.

6. S. Sharmin and D. Chakma, "Attention-based convolutional neural network for bangla sentiment analysis," *Ai & Society*, vol. 36, no. 1, pp. 381–396, 2021. Accessed: 2022-07-10.

7. S. Azmin and K. Dhar, "Emotion detection from bangla text corpus using naive bayes classifier," in *2019 4th International Conference on Electrical Information and Communication Technology (EICT)*, pp. 1–5, IEEE, 2019. Accessed: 2022-07-10.

8. N. Banik and M. H. H. Rahman, "Evaluation of naïve bayes and support vector machines on bangla textual movie reviews," in *2018 international conference on Bangla speech and language processing (ICBSLP)*, pp. 1–6, IEEE, 2018. Accessed: 2022-07-10.

9. M. Khan and K. Malik, "Sentiment classification of customers reviews about automobiles in roman urdu," in *Future of Information and Communication Conference*, pp. 630–640, Springer, 2018. Accessed: 2022-07-10.

10. K. I. Islam, M. S. Islam, and M. R. Amin, "Sentiment analysis in bengali via transfer learning using multi-lingual BERT," *CoRR*, vol. abs/2012.07538, 2020. Accessed: 2022-07-12.

11. S. Khan and M. Shahid, "Hindi/bengali sentiment analysis using transfer learning and joint dual input learning with self attention," *arXiv preprint arXiv:2202.05457*, 2022. Accessed: 2022-07-12.

12. S. Sarker, "Banglabert: Bengali mask language model for bengali language understading," 2020. Accessed: 2022-07-10.

13. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017. Accessed: 2022-07-10.

14. S. Sazzed, "Cross-lingual sentiment classification in low-resource bengali language," in *Proceedings of the Sixth Workshop on Noisy User-generated Text (W-NUT 2020)*, pp. 50–60, 2020. Accessed: 2022-07-10.

15. KhondokerIslam, "Bengali Sentiment." https://github.com/KhondokerIslam/Bengali_Sentiment/tree/master/Dataset, 2020. Accessed: 2022-07-10.

16. E. Hossain, "Bengali Restaurent Reviews." https://github.com/eftekhar-hossain/Bengali-Restaurant-Reviews, 2020. Accessed: 2022-07-10.

17. E. Hossain, "Bengali Book Reviews." https://github.com/eftekhar-hossain/Bengali-Book-Reviews, 2020. Accessed: 2022-07-10.

18. A. Rahman, "Bangla_ABSA_Datasets/Cricket." https://github.com/atik-05/Bangla_ABSA_Datasets/blob/master/Cricket.xlsx, 2018. Accessed: 2022-07-10.

19. A. Rahman, "Bangla_ABSA_Datasets/Restaurant." https://github.com/atik-05/Bangla_ABSA_Datasets/blob/master/Restaurant.xlsx, 2018. Accessed: 2022-07-10.

20. Ashik and M. A. U. Zaman, "Data set for sentiment analysis on bengali news comments," 2019. Accessed: 2022-07-10.

21. Y. Wu, M. Schuster, Z. Chen, Q. V. Le, M. Norouzi, W. Macherey, M. Krikun, Y. Cao, Q. Gao, K. Macherey, J. Klingner, A. Shah, M. Johnson, X. Liu, L. Kaiser, S. Gouws, Y. Kato, T. Kudo, H. Kazawa, K. Stevens, G. Kurian, N. Patil, W. Wang, C. Young, J. Smith, J. Riesa, A. Rudnick, O. Vinyals, G. Corrado, M. Hughes, and J. Dean, "Google's neural machine translation system: Bridging the gap between human and machine translation," *CoRR*, vol. abs/1609.08144, 2016. Accessed: 2022-07-12.